

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:55:45 ; Search time 29 Seconds
(without alignments)
113.681 Million cell updates/sec

Title: US-10-006-223-1

Perfect score: 82
Sequence: 1 VVGGRAAQGEPPFMV 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

1 number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rhodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	93.9	259	Q54168	streptomyce
2	62	75.6	255	Q9Y7A9	metarhizium
3	62	75.6	261	Q00344	cochliobolu
4	61	74.4	311	Q8WZM5	trichoderma
5	60	73.2	254	Q01136	metarhizium
6	60	73.2	256	Q9Y842	metarhizium
7	59	72.0	265	Q74696	phaeosphaer
8	56	68.3	360	Q9W1X6	drosophila
9	55	67.1	281	Q76898	drosophila
10	54	65.9	242	Q96897	lumbricus b
11	54	65.9	251	Q9VXC9	drosophila
12	54	65.9	294	Q9V7R3	drosophila
13	54	65.9	974	Q90WD8	bufo japoni
14	53	64.6	288	Q46151	pacifastacu
15	53	64.6	855	Q9JJI7	rattus norv
16	52	63.4	155	Q95VT4	homarus ame

17	52	63.4	548	16	Q9KRJ1	Q9krj1 vibrio chol
18	51	62.2	262	5	Q9VRF9	Q9vrf9 drosophila
19	51	62.2	405	5	Q44331	Q44331 manduca sex
20	51	62.2	422	4	Q8WVC1	Q8wvc1 homo sapien
21	51	62.2	638	11	Q8R0P5	Q8r0p5 mus musculu
22	50	61.0	243	5	Q01310	Q01310 botryllus s
23	50	61.0	271	5	Q9VRS6	Q9vrs6 drosophila
24	50	61.0	281	5	Q8SYB5	Q8syb5 drosophila
25	50	61.0	1379	5	Q9V4N6	Q9v4n6 drosophila
26	49	59.8	272	5	Q9XYV6	Q9xyv6 rhizopertha
27	49	59.8	275	6	Q28609	Q28609 oryctolagus
28	49	59.8	286	5	Q96900	Q96900 scolopendra
29	49	59.8	542	5	Q917J3	Q917j3 drosophila
30	49	59.8	613	13	Q03711	Q03711 xenopus lae
31	49	59.8	845	13	Q9DGR1	Q9dgr1 xenopus lae
32	48	58.5	25	5	P81802	P81802 eisenia foe
33	48	58.5	213	5	Q9VEN0	Q9ven0 drosophila
34	48	58.5	265	5	Q9VVT3	Q9vvt3 drosophila
35	48	58.5	433	5	Q8T3A2	Q8t3a2 ciona intes
36	47	57.3	223	5	Q9XY58	Q9xy58 ctenocephal
37	47	57.3	232	5	Q9TY16	Q9ty16 penaeus van
38	47	57.3	263	5	Q9TY16	Q9ty16 penaeus van
39	47	57.3	266	5	Q27761	Q27761 penaeus van
40	47	57.3	277	5	Q96899	Q96899 scolopendra
41	47	57.3	279	11	Q9QZ74	Q9qz74 rattus norv
42	47	57.3	377	6	P79343	P79343 bos taurus
43	47	57.3	417	11	Q8VHJ4	Q8vhj4 rattus norv
44	47	57.3	819	16	Q9A500	Q9a500 caulobacter
45	46	56.1	40	6	Q9TRU3	Q9tru3 oryctolagus

ALIGNMENTS

RESULT 1

Q54168 ID Q54168 PRELIMINARY; PRT; 259 AA.
AC Q54168;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trypsinogen.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14544;
RA Katoh T., Kikuchi N., Nagata K., Yoshida N.;
RT "Cloning and expression of Trypsin-Like enzyme from Streptomyces
RT fradiae for comparative analysis of functional regions of Streptomyces
RT and Mammalian Trypsins.";
RL J. Ferment. Bioeng. 80:440-445(1995).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; D16687; BAA04089.1; --
DR HSSP; P00775; 1SGT.
DR MEROPS; S01.101; --
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT CHAIN 39 259 TRYPSIN.
SQ SEQUENCE 259 AA; 26702 MW; 8B6C8496AB840FCF CRC64;
Query Match 93.9%; Score 77; DB 2; Length 259;
Best Local Similarity 93.8%; Pred. No. 1.8e-05;

```
Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
   ||||| ||||| |||||
Db 39 VVGGTAAQGEFFPMV 54

RESULT 2
ID Q9Y7A9 PRELIMINARY; PRT; 255 AA.
AC Q9Y7A9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
Trypsin-related protease.
TRY2.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
NCBI_TaxID=5530;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ME1;
RA Screen S.E., St Leger R.J.;
RT "Isolation of multiple protease genes from the entomopathogenic fungus
RL Metarhizium anisopliae."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF130865; AAD29675.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.103; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 255 AA; 26289 MW; 88DD979ED300E4B7 CRC64;

Query Match 75.6%; Score 62; DB 3; Length 255;
Best Local Similarity 68.8%; Pred. No. 0.0078;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
   ||||| ||||| |||||
Db 30 IVGGEAAQGEFFYIV 45

RESULT 3
ID Q00344 PRELIMINARY; PRT; 261 AA.
AC Q00344;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ALP1.
GN ALP1.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
NCBI_TaxID=5017;
[1]
SEQUENCE FROM N.A.
RC STRAIN=SB11;
RX MEDLINE=96212997; PubMed=8634479;
RA Murphy J.M., Walton J.D.;
RT "Three extracellular proteases from Cochliobolus carbonum: cloning and
RT targeted disruption of ALP1."
Mol. Plant Microbe Interact. 9:290-297(1996).
GN
```

```
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; U39500; AAB03851.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.103; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 261 AA; 26052 MW; FD4B4A02BF44BDE1 CRC64;

Query Match 75.6%; Score 62; DB 3; Length 261;
Best Local Similarity 68.8%; Pred. No. 0.008;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
   ||||| ||||| |||||
Db 31 IVGTTAAAGEFFIV 46

RESULT 4
ID Q8WZM5 PRELIMINARY; PRT; 311 AA.
AC Q8WZM5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
Protease P27 (Fragment).
GN P27.
OS Trichoderma harzianum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Trichoderma.
NCBI_TaxID=5544;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CECT 2413;
RA Suarez M.B.;
RL Thesis (1999), University of Salamanca, Salamanca, SPAIN.
DR EMBL; AJ249721; CAC80694.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser.protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 311 AA; 31200 MW; B31CC1493DB2263F CRC64;

Query Match 74.4%; Score 61; DB 3; Length 311;
Best Local Similarity 68.8%; Pred. No. 0.015;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
   ||||| ||||| |||||
Db 83 IVGTTAALGEFFIV 98

RESULT 5
ID Q01136 PRELIMINARY; PRT; 254 AA.
AC Q01136;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trypsin-like protease 1 precursor.
GN TYR1.
```

```

FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 30 256 TRYPsin-RELATED PROTEASE.
SQ SEQUENCE 256 AA; 26201 MW; 34696608745CB982 CRC64;

Query Match 73.2%; Score 60; DB 3; Length 256;
Best Local Similarity 68.8%; Pred. No. 0.018;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGFFPMV 16
Db 30 IVGGSPAAGEFFPIV 45
:||||:|||||:
:||||:|||||:

RESULT 7
O74696 PRELIMINARY; PRT; 265 AA.
ID O74696
AC O74696;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trypsin-like protease.
GN SNPI
OS Phaeosphaeria nodorum (Septoria nodorum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Phaeosphaeriaceae; Phaeosphaeria.
OC NCBI_TaxID=13684;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LAW95;
RX MEDLINE=20255037; PubMed=10796020;
RA Carlie A.J., Bindschedler L.V., Bailey A.M., Bowyer P.,
RA Clarkson J.M., Cooper R.M.;
RT "Characterization of SNPI, a cell wall-degrading trypsin, produced
RT during infection by Stagonospora nodorum.";
RL Mol. Plant Microbe Interact. 13:538-550(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF092435; AAC61777.1; -.
DR HSP; P35049; ITRY.
DR MEROPS; S01.103; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 265 AA; 26296 MW; 3B9DB00640DA4FEA CRC64;

Query Match 72.0%; Score 59; DB 3; Length 265;
Best Local Similarity 62.5%; Pred. No. 0.028;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGFFPMV 16
Db 37 IVGGTTASAGDFFPIV 52
:||||:|||||:
:||||:|||||:

RESULT 8
Q9WLX6 PRELIMINARY; PRT; 360 AA.
ID Q9WLX6
AC Q9WLX6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG3700 protein (GH06673p).
GN CG3700.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pharyngota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fouts C., Gabrieliata C.C., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasakman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.S.;
 RA "The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AE003459; AAF46926.1; -;
 DR ENBL; AY069055; AAL39200.1; -;
 DR HSP; P20160; 1A7S.
 DR FlyBase; FBgn0034796; CG3700.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 360 AA; 39809 MW; 619EBB215A182023 CRC64;
 SQ
 Query Match 68.3%; Score 56; DB 5; Length 360;
 Best Local Similarity 66.7%; Pred. No. 0.13;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease.
 KW SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;
 SQ
 Query Match 67.1%; Score 55; DB 5; Length 281;
 Best Local Similarity 62.5%; Pred. No. 0.15;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 DR EMBL; AL031027; CAA19843.2; -;
 DR HSP; P00756; 1CHG.
 DR FlyBase; FBgn0025385; EG-80H7.1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRY

RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:"	
RI	"the genome sequence of <i>Drosophila melanogaster</i> ;"	
RL	Science 287:2185-2195(2000).	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE	
CC	TRYPSIN FAMILY.	
CC	EMBL; AE003503; AAF48647.1; -	
DR	HSP; P00761; 1EPT.	
DR	FlyBase; FBN0030774; CG9675.	
DR	InterPro; IPR001314; Chymotrypsin.	
DR	InterPro; IPR001254; Ser protease_Try.	
DR	Pfam; PF00089; trypsin; 1.	
DR	PRINTS; PR00722; CHYMOTRYPSIN.	
DR	SMART; SW00020; Tryp_SPC; 1.	
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.	
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.	
KW	Hydrolase; Serine protease.	
SQ	SEQUENCE 251 AA; 26666 MW; 2248D38D6CDFD809 CRC64;	
Query Match 65.9%; Score 54; DB 5; Length 251;		
Best Local Similarity 50.0%; Pred. No. 0.2;		
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps		
Qy	1 VVGGTAAQGEFFPMV 16	
	: : : :	
Db	28 IVGGTKAREGFPHQI 43	
RESULT 12		
Q9V7R3	PRELIMINARY;	
ID	Q9V7R3	PRT; 294 AA.
AC	Q9V7R3;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	C4927 protein.	
GN	C4927	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Prerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
EN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RC	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,	
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	

```
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR ENBL; AE003806; AAF57983.1; -.
DR HSP; P20160; IA7S.
DR FlyBase; FBgn0034139; CG4927.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 294 AA; 32482 MW; 4867A6C355BB3C51 CRC64;

Query Match 65.9%; Score 54; DB 5; Length 294;
Best Local Similarity 66.7%; Pred. No. 0.23;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPM 15
DB 37 IVGGAKAAGREFFPM 51

RESULT 13
Q90WD8 Q90WD8 PRELIMINARY; PRT; 974 AA.
ID Q90WD8;
RA TISSUE=OVIDUCTAL PARS RECTA;
RA Hiyoishi M., Takamune K., Mita K., Kubo H., Sugimoto Y., Katagiri C.;
RT "Oviductin, the oviductal protease that mediates gamete interaction by
RT affecting the vitelline envelope in Bufo japonicus: Its molecular
RT cloning and analyses of expression and post-translational
RT activation.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB070367; BAB63372.1; -.
DR MEROPS; S01.240; -.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00089; trypsin; 2.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR Hydrolase; Serine protease.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 974 AA; 107647 MW; F19705A470465553 CRC64;
```

```
Query Match 65.9%; Score 54; DB 13; Length 974;
Best Local Similarity 62.5%; Pred. No. 0.85;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
DB 50 IVGTSATKVGESPMV 65

RESULT 14
Q46151 Q46151 PRELIMINARY; PRT; 268 AA.
ID Q46151;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4) (Fragment).
OS Pacifastacus leniusculus (Signal crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Astacidae; Pacifastacus.
OX NCBI_TaxID=6720;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEPATOPANCREAS;
RX MEDLINE=9281768; PubMed=10355637;
RA Hernandez-Cortes M.P., Cerenius L., Garcia-Carreno F.L., Soderhall K.;
RT "Trypsin from Pacifastacus leniusculus Hepatopancreas: Purification
RT and cDNA Cloning of the Synthesized Zymogen.";
RL Biol. Chem. 380:499-501(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ222658; CAA10915.1; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.122; -.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Signal.
KW NON_TER 1
FT SIGNAL <1 17 POTENTIAL.
FT CHAIN 31 266 TRYPSIN.
SQ SEQUENCE 268 AA; 28482 MW; 9080F9736412FEF5 CRC64;

Query Match 64.6%; Score 53; DB 5; Length 268;
Best Local Similarity 56.2%; Pred. No. 0.32;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
DB 32 IVGTDASLGEFFYQL 47

RESULT 15
Q9JJI7 Q9JJI7 PRELIMINARY; PRT; 855 AA.
ID Q9JJI7;
AC Q9JJI7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Membrane bound serine protease (Membrane bound arginine specific
DE serine protease).
GN MESP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tezuka S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=DUODENUM;
RA Inoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AB037898; BAB03502.1; -;
DR EMBL; AB049189; BAB13765.1; -;
DR HSSP; P00763; LDPO.
DR MEROPS; S01.302; -;
DR InterPro; IPR002106; AACRNA_ligaseII.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
FT VARIANT 665 665
SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECF6CF03D CRC64;

Query Match 64.6%; Score 53; DB 11; Length 855;
Best Local Similarity 62.5%; Pred. No. 1.1;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 16
|||||:|:|:
Db 615 VVGTTNADGEWPMQV 630

rch completed: April 11, 2003, 17:58:07
time : 31 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:55:20 ; Search time 11 Seconds
(without alignments)
60.329 Million cell updates/sec

Title: US-10-006-223-1

Perfect score: 82

Sequence: 1 VVGGRRAQGEFFFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

1 number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	259	1 TRYP_STRGR	P00775 streptomyce
2	62	75.6	268	1 TRYP_STRGA	Q54179 streptomyce
3	59	72.0	248	1 TRYP_FUSOX	P35049 fusarium ox
4	58	70.7	20	1 TRYL_STREX	P80420 streptomyce
5	53	64.6	855	1 STI4_MOUSE	P56677 mus musculus
6	52	63.4	216	1 CTR2_VESOR	P00768 vespa orien
7	52	63.4	237	1 TRYP_ASTFL	P00765 astacus flu
8	52	63.4	254	1 CTRL_HALRU	P35003 haliois ru
9	52	63.4	274	1 FA9_SHEEP	P16291 ovis aries
10	51	62.2	638	1 KAL_MOUSE	P26262 mus musculus
11	51	62.2	855	1 STI4_HUMAN	Q9Y5Y6 homo sapien
12	50	61.0	32	1 TRYP_PENMO	P35050 penaeus mon
13	50	61.0	218	1 CTR2_VESCR	P00769 vespa crabr
14	49	59.8	219	1 CAP7_PIG	P80015 sus scrofa
15	49	59.8	431	1 ACRO_RABIT	P48038 oryctolagus
16	49	59.8	625	1 FALL_HUMAN	P03951 homo sapien
17	48	58.5	20	1 COG3_PARCM	P20734 paralithode
18	48	58.5	271	1 S24D_ANOGA	Q17004 anopheles g
19	48	58.5	638	1 KAL_HUMAN	P03952 homo sapien
20	48	58.5	638	1 KAL_RAT	P14272 rattus norv
21	47	57.3	60	1 ACRO_CAPIH	P10626 capra hircu
22	47	57.3	416	1 FA9_BOVIN	P00741 bos taurus
23	46	56.1	271	1 FA9_PIG	P16293 sus scrofa
24	46	56.1	275	1 FA9_RABIT	P16292 oryctolagus
25	46	56.1	705	1 CIR_HUMAN	P00736 homo sapien
26	45	54.9	20	1 COG2_CHIOP	P34154 chionocete
27	45	54.9	20	1 COG3_CHIOP	P34155 chionocete
28	45	54.9	20	1 COG4_PARCM	P20732 paralithode
29	45	54.9	20	1 COG5_PARCM	P20733 paralithode
30	45	54.9	267	1 ELNE_HUMAN	P08246 homo sapien
31	45	54.9	285	1 FA9_CAVPO	P16295 cavia porce
32	45	54.9	418	1 HATT_HUMAN	O60235 homo sapien
33	45	54.9	422	1 DES1_HUMAN	Q9U152 homo sapien

RESULT 1

TRYP_STRGR

ID TRYP_STRGR STANDARD; PRT; 259 AA.

AC P00775; 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Trypsin precursor (EC 3.4.21.4) (SGT).

GN SPRT.

OS Streptomyces griseus.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1911;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 10137;

RX MEDLINE=92095977; PubMed=1755852;

RA Kim J.C., Cha S.H., Jeong S.T., Oh S.K., Byun S.M.;

RT "Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin gene."

RL Biochem. Biophys. Res. Commun. 181:707-713(1991).

RN [2]

RP SEQUENCE OF 37-259.

RX MEDLINE=75127940; PubMed=804314;

RA Olafson R.W., Jurasek L., Carpenter M.R., Smillie L.B.;

RT "Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragments and complete sequence."

RL Biochemistry 14:1168-1177(1975).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).

RX MEDLINE=88286735; PubMed=3135412;

RA Read R.J., James M.N.G.;

RT "Refined crystal structure of Streptomyces griseus trypsin at 1.7-A resolution."

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M64471; AAA26820.1; ALT_SEQ.

DR PIR; A00962; TRSWG

DR PIR; JQ1302; JQ1302.

DR PDB; ISGT; 16-JUL-88.

DR MEROPS; S01.101; .

DR InterPro; IPR001254; Chymotrypsin.

DR Pfam; PF00089; trypsin, I.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp_Spc; 1.

```
DR PROSITE; PS50240; TRYP SIN DOM; 1.
DR PROSITE; PS00134; TRYP SIN HIS; 1.
DR PROSITE; PS00135; TRYP SIN SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal; 3D-structure.
FT SIGNAL 1 32
FT PROPEP 33 36 ACTIVATION PEPTIDE.
FT CHAIN 37 259 TRYP SIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM.
FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
FT ACT_SITE 208 208 CHARGE RELAY SYSTEM.
FT DISULFID 58 74
FT DISULFID 177 192
FT DISULFID 204 233
FT SITE 202 202 REQUIRED FOR SPECIFICITY.
FT CONFLICT 95 96 MISSING (IN REF. 2).
FT STRAND 38 38
FT TURN 39 39
FT STRAND 41 42
FT TURN 45 46
FT TURN 49 50
FT STRAND 51 54
FT TURN 55 57
FT STRAND 58 64
FT TURN 65 66
FT STRAND 67 70
FT HELIX 72 74
FT STRAND 79 80
FT STRAND 85 88
FT STRAND 92 92
FT TURN 93 94
FT TURN 96 97
FT STRAND 99 108
FT TURN 110 111
FT STRAND 120 124
FT STRAND 134 135
FT TURN 140 141
FT STRAND 145 150
FT TURN 156 157
FT STRAND 163 163
FT STRAND 165 172
FT HELIX 174 184
FT TURN 187 189
FT STRAND 190 193
FT TURN 196 198
FT STRAND 202 202
FT TURN 205 206
FT TURN 208 209
FT STRAND 211 215
FT TURN 217 218
FT STRAND 221 229
FT TURN 236 237
FT STRAND 240 244
FT HELIX 245 257
FT TURN 258 258
SQ SEQUENCE 259 AA; 26776 MW; 050233AFF1F64823 CRC64;

Query Match 100.0%; Score 82; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
DB 37 VVGGTAAAGGEPFPMV 52

RESULT 2
ID TRYP STRGA STANDARD; PRT; 268 AA.
AC Q54179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin-like protease precursor (EC 3.4.21.-).

OS Streptomyces glaucescens.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1907;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / ETH 22794 / GLA.0;
RA Hintemann G.;
RT "Cloning, partial characterization and nucleotide sequence of a
RT trypsin-like protease gene from Streptomyces glaucescens.";
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U13770; AAA21336.1; -.
CC HSP; P00775; LSGT.
CC MEROPS; S01.101; -.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001354; Ser. protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS00240; TRYP SIN DOM; 1.
CC PROSITE; PS00134; TRYP SIN HIS; 1.
CC PROSITE; PS00135; TRYP SIN SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 41 POTENTIAL.
FT PROPEP 42 45 ACTIVATION PEPTIDE.
FT CHAIN 46 268 TRYP SIN-LIKE PROTEASE.
FT ACT_SITE 82 82 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 127 127 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 67 83 BY SIMILARITY.
FT DISULFID 187 202 BY SIMILARITY.
FT DISULFID 213 242 BY SIMILARITY.
FT SITE 211 211 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 268 AA; 27550 MW; 01D05E0A4D5D177A CRC64;

Query Match 75.6%; Score 62; DB 1; Length 268;
Best Local Similarity 75.0%; Pred. No. 0.0011;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
DB 46 VVGGTAAAGGEPFPMV 61

RESULT 3
ID TRYP FUSOX STANDARD; PRT; 248 AA.
AC P35049;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RC MEDLINE=93324511; PubMed=8332590;
RA Rypniewski W.R., Hastrup S., Betzel C., Dauter M., Dauter Z.,
RA Papendorf G., Branner S., Wilson K.S.;
RT "The sequence and X-ray structure of the trypsin from Fusarium
```

oxytumorum.";
 RL Protein Eng. 6:341-348(1993).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).
 RA Rypniewski W.R., Dammann C., von der Osten C., Dauter M.,
 RA Wilson K.S.;
 RT "Structure of inhibited trypsin from *Fusarium oxysporum* at 1.55 A.";
 RL Acta Crystallogr. D 51:73-84(1995).
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 EMBL; S63827; AAB27568.1; -;
 PDB; 1TRY; 01-JAN-96.
 DR MEROPS; S01.103; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYPSIN; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydroxylase; Serine protease; Zymogen; Signal; 3D-structure.
 FT SIGNAL 1 17
 FT PROPEP 18 25
 FT CHAIN 26 248
 FT ACT_SITE 65 65
 FT ACT_SITE 108 108
 FT ACT_SITE 204 204
 FT DISULFID 50 66
 FT DISULFID 174 189
 FT DISULFID 200 225
 FT SITE 198 198
 SQ SEQUENCE 248 AA; 24576 MW; 1A0EBA88C3E70294 CRC64;
 Query Match 72.0%; Score 59; DB 1; Length 248;
 Best Local Similarity 62.5%; Pred. No. 0.0034;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 1 VVGTTAAAGCEPFFMV 16
 :|||||:|:|:|:
 25 IVGTSASAGDPFFIV 40
 ULT 4
 STREX STANDARD; PRT; 20 AA.
 AC P80420;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trypsin-like protease (SC 3.4.21.-) (fragment).
 OS Streptomyces exfoliatus (Streptomyces hydrogenans),
 OC Bacteria; Actinobacteria; Actinobacteria (Class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1905;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=SMF13;
 RX MEDLINE=95291424; PubMed=7773379;
 RA Kim I.S., Lee K.J.;
 RT "Physiological roles of leupeptin and extracellular proteases in
 RT mycelium development of *Streptomyces exfoliatus* SMF13.";
 RL Microbiology 141:1017-1025(1995).

CC -!- FUNCTION: INVOLVED IN MYCELIUM DIFFERENTIATION.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2A.
 DR MEROPS; S01.101; -;
 DR InterPro; IPR001254; Ser_protease_Try.
 DR PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
 DR PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
 KW Hydroxylase; Serine protease.
 FT NON_TER 20
 SQ SEQUENCE 20 AA; 2129 MW; 4568F56D0E7393AF CRC64;
 Query Match 70.7%; Score 58; DB 1; Length 20;
 Best Local Similarity 84.6%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 VGGTAAAGCEPFF 14
 :|||||:|:|:|:
 Db 2 VGGTAAAGCNFFP 14
 RESULT 5
 ST14_MOUSE
 ID ST14_MOUSE STANDARD; PRT; 855 AA.
 AC P56677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Suppressor of tumorigenicity 14 (SC 3.4.21.-) (Epithin).
 GN ST14 OR PRSS14.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B.175CID; TISSUE=Thymus;
 RX MEDLINE=99216440; PubMed=10199918;
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
 RA Schwartz R.H.;
 RT "Cloning and chromosomal mapping of a gene isolated from thymic
 RT stromal cells encoding a new mouse type II membrane serine protease,
 RT epithin, containing four LDL receptor modules and two CUB domains.";
 RL Immunogenetics 49:420-428(1999).
 RN [2]
 RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
 RC STRAIN=C.B.175CID; TISSUE=Thymus;
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
 CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
 CC TESTIS, AND BRAIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 EMBL; AF042822; AAD02230.3; -;
 DR EMBL; BC005496; AAH05496.1; -;
 DR HSP; P20231; IAAO.
 DR MEROPS; S01.302; -;
 DR MGD; MGI:1338881; St14.

```

RT Orientalis.";
RL Blochm. Biophys. Acta 668:197-200(1981).
CC -|- CATALYTIC ACTIVITY; Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -|- SUBCELLULAR LOCATION; Extracellular.
CC -|- SIMILARITY; BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A00954; KTVH20.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT ACT_SITE 39 39 CHARGE RELAY SYSTEM.
FT ACT_SITE 82 82 CHARGE RELAY SYSTEM.
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.
FT DISULFID 25 40
FT DISULFID 146 159
FT DISULFID 169 193
SQ SEQUENCE 216 AA; 23471 MW; F235BF992AEFDE1 CRC64;

Query Match 63.4%; Score 52; DB 1; Length 216;
Best Local Similarity 50.0%; Pred. No. 0.052;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps

Qy 1 VVGTRAAQGSEFFPMV 16
:|||||:|:|:|
Db 1 IVGGTNAPRGKYPQV 16

RESULT 7
TRYP ASTFL STANDARD; PRT; 237 AA.
ID TRYP ASTFL STANDARD; PRT; 237 AA.
AC P00765;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin I (EC 3.4.21.4).
OS Astacus fluviatilis (Broad-fingered crayfish) (Astacus astacus).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Astacidea; Astacoidea; Astacidae; Astacus.
ON NCBI_taxid=6715;
RX [1]
RP SEQUENCE.
RX MEDLINE=83178967; PubMed=6838862;
RA Titani K., Sasagawa T., Woodbury R.G., Ericsson L.H., Doream H.,
RA Kraemer M., Neureath H., Zwilling R.;
RT "Amino acid sequence of crayfish (Astacus fluviatilis) trypsin if.";
RL Biochemistry 22:1459-1465(1983).
CC -|- CATALYTIC ACTIVITY; Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION; Extracellular.
CC -|- MISCELLANEOUS; TRYPSIN I IS ONE OF FIVE FORMS OF THE ENZYME KNOWN
CC TO BE PRESENT IN CRAYFISH. THIS PROTEIN IS MORE ACIDIC THAN
CC MAMMALIAN TRYPSIN.
CC -|- SIMILARITY; BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A00951; TRCY1.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.122; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas.

```

```

FT ACT_SITE 45 45 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 46
FT DISULFID 159 174
FT DISULFID 185 213
FT SITE 183 183
SQ SEQUENCE 237 AA; 25021 MW; 4072133E55022C76 CRC64;

Query Match
Best Local Similarity. 63.4%; Score 52; DB 1; Length 237;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVGTRAAOGEPFPMV 16
DB 1 IVGTDVGLGEFFYQL 16

RESULT 8
CTRL_HALRU STANDARD; PRT; 254 AA.
AC F35003;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DT Chymotrypsin-like serine proteinase precursor (EC 3.4.21.-).
OC Haliotis rufescens (California red abalone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;
OC Haliotidae; Haliotis.
OX NCBI_TaxID=6454;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Intestine;
RX MEDLINE=93343624; PubMed=8242947;
RA Groppe J.C., Morse D.B.;
RT "Molluscan chymotrypsin-like protease: structure, localization, and
RT substrate specificity.";
RJ Arch. Biochem. Biophys. 305:159-169(1993).
CC -!- FUNCTION: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -!- ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: EXTRACELLULAR.
CC -!- SUBCELLULAR LOCATION: OF THE INTESTINE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X71438; CAA50572.1; --
CC PIR; S35585; S35585.
CC PIR; S32750; S32750.
CC HSP; P00763; LDPO.
CC
CC MEROPS; S01.121; --
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Signal; Zymogen.
CC SIGNAL 1 18
CC PROPEP 19 23
CC CHAIN 24 254
CC ACT_SITE 68 68
CC ACT_SITE 117 117

```

```

FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 53 69 BY SIMILARITY.
FT DISULFID 146 218 BY SIMILARITY.
FT DISULFID 181 199 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
SQ SEQUENCE 254 AA; 27250 MW; ADAA9A8A22BEFCEC CRC64;

Query Match
Best Local Similarity. 63.4%; Score 52; DB 1; Length 254;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAOGEPF 14
DB 24 IVGGSNAARAGEFW 37

RESULT 9
FA9_SHEEP
ID FA9_SHEEP STANDARD; PRT; 274 AA.
AC P16291;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor IX (EC 3.4.21.22) (Christmas factor) (Fragment).
GN F9.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90152675; PubMed=2303254;
RA Sarkar G., Koester D.D., Sommer S.S.;
RT "Direct sequencing of the activation peptide and the catalytic domain
RT of the factor IX gene in six species.";
RJ Genomics 6:133-143(1990).
CC -!- FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT
CC PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY
CC CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++
CC IONS, PHOSPHOLIPIDS, AND FACTOR VIIIa.
CC -!- CATALYTIC ACTIVITY: Hydrolyzes one Arg-|-ile bond in factor X to
CC form factor Xa.
CC -!- SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIA, WHICH EXCISES THE
CC ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2
CC CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M26233; AAA31520.1; --
CC HSP; P16293; IPEX.
CC
CC MEROPS; S01.214; --
CC InterPro; IPR001254; Ser_protease_Try.
CC InterPro; IPR000294; VitK_dep_GLA.
CC Pfam; PF00089; trypsin; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00011; GLU_CARBOXYLATION; PARTIAL.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Blood coagulation; Plasma; Serine protease; Calcium-binding;
CC Hydrolase; Glycoprotein.
CC NON_TER 1 1
CC ACT_SITE 89 89 CHARGE RELAY SYSTEM.
CC ACT_SITE 137 137 CHARGE RELAY SYSTEM.
CC ACT_SITE 233 233 CHARGE RELAY SYSTEM.

```

FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 274 274
SQ SEQUENCE 274 AA; 30595 MW; D3617FC3BD33E9B CRC64;

Query Match 63.4%; Score 52; DB 1; Length 274;
Best Local Similarity 62.5%; Pred. No. 0.067; 3; Mismatches 0; Gaps 0;
Matches 10; Conservative 3; Indels 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
49 VVGEDAAQGEFFPMV 64
RESULT 10
KAL MOUSE
ID KAL_MOUSE STANDARD; PRT; 638 AA.
P26262;
01-MAY-1992 (Rel. 22, Created)
01-MAY-1992 (Rel. 22, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
DE Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)
DE (Kininogenin) (Fletcher factor).
GN KLK1 OR KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
STRAIN=BALB/c; TISSUE=Liver;
MEDLINE=91090844; PubMed=2264928;
RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.,
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa
CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
CC kininogen to release bradykinin.
CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL; M58588; AAA63393.1; .
CC PIR; A36557; KQMSPL.
CC HSSP; P00750; 1RTF.
CC MEROPS; S01.212; .
CC MGD; MGI:102849; KIK3.
CC InterPro; IPR000177; Apple.
CC InterPro; IPR003314; Chymotrypsin.
CC InterPro; IPR003014; PAN.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00024; PAN; 4.

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 O-LINKED (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 62.2%; Score 51; DB 1; Length 638;
Best Local Similarity 56.2%; Pred. No. 0.24;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
391 IVGGTNASLGEWPMQV 406
RESULT 11
ST14_HUMAN
ID ST14_HUMAN STANDARD; PRT; 855 AA.
AC Q9Y5Y6; Q9H3S0; Q9HCA3; Q9BS01; Q9HB36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matritypase) (Membrane-
DE type serine protease 1) (MT-SPL) (Proctamin) (Serine protease TAGD-15)
DE (Tumor associated differentially-expressed gene-15 protein).
GN ST14 OR PRSS14 OR SMC19 OR TAGD15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99030581; PubMed=10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
 RT protease with trypsin-like activity";
 RN J. Biol. Chem. 274:18231-18236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432178; PubMed=10500122;
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 RT dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue";
 RN Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Prostate;
 RA Yamaguchi N., Mitsui S.;
 RT "Molecular cloning of a novel transmembrane serine protease expressed
 RT in human prostate";
 RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
 RA O'Brien T.J.;
 RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 327-855 FROM N.A.
 RX TISSUE=Muscle;
 RA Strausberg R.;
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 340-664 FROM N.A.
 RX Cao J., Fan W., Zheng S.;
 RA "Genomic analysis of a novel human serine protease SN19";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP CHARACTERIZATION.
 RX TISSUE=Milk;
 RA MEDLINE=99303582; PubMed=10373425;
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
 RT "Purification and characterization of a complex containing matriptase
 RT and a Kunitz-type serine protease inhibitor from human milk";
 RN J. Biol. Chem. 274:18237-18242(1999).
 CC -!- FUNCTION: DEGRADES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
 CC OR LYS AS THE P1 SITE.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF118224; RAD42765.2; -;
 CC EMBL; AF133086; AAF00109.1; -;
 CC EMBL; AB030036; BAB20376.1; -;
 CC EMBL; AF057145; AAG15395.1; -;
 CC EMBL; BC005826; AAH05826.1; -;
 CC EMBL; AF283256; AAG13949.1; -;
 CC HSSP; P00763; IDPO. ST14.
 CC Genem; HGNC:11344; ST14.
 CC MIM; 606797; -;
 CC MEROPS; S01.302; -;
 CC InterPro; IPR000859; CUB domain.
 CC InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR002172; LDL recept A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00431; CUB; 2.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLa; 3.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55
 FT TRANSMEM 56 76
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT CYTOPLASMIC (POTENTIAL).
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT CUB 1.
 FT CUB 2.
 FT LDL-RECEPTOR CLASS A 1.
 FT LDL-RECEPTOR CLASS A 2.
 FT LDL-RECEPTOR CLASS A 3.
 FT LDL-RECEPTOR CLASS A 4.
 FT SERINE PROTEASE.
 FT ACT_SITE 554 554
 FT ACT_SITE 556 556
 FT ACT_SITE 711 711
 FT ACT_SITE 805 805
 FT CARBOHYD 109 109
 FT CARBOHYD 302 302
 FT CARBOHYD 485 485
 FT CARBOHYD 772 772
 FT CONFLICT 327 329
 FT CONFLICT 381 381
 FT CONFLICT 674 674
 FT SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
 Query Match 62.2%; Score 51; DB 1; Length 855;
 Best Local Similarity 62.5%; Pred. No. 0.33;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVGTTAAQGEPPFMY 16
 Db 615 VVGTTAAQGEPPFMY 630
 RESULT 12
 TRYP_PENMO STANDARD; PRT; 32 AA.
 AC P35050;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trypsin (EC 3.4.21.4) (Fragment).
 OS Penaeus monodon (Penaeid shrimp).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata;
 OC Penaeoidea; Penaeidae; Penaeus.
 OX NCBI_TaxID=6687;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Midgut;
 RX MEDLINE=91151586; PubMed=1963309;
 RA Lu P.J., Liu H.C., Tsai I.H.;
 RT "The midgut trypsins of shrimp (Penaeus monodon). High efficiency
 RT toward native protein substrates including collagens";
 RL Biol. Chem. Hoppe-Seyler 371:851-859(1990).
 CC -!- FUNCTION: HIGH EFFICIENCY TOWARD NATIVE PROTEIN SUBSTRATES
 CC INCLUDING COLLAGENS. CC


```
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; S11537; S11537.
DR MEROPS; S01.111; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR PFAM; PF00089; trypsin; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; PARTIAL.
DR PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
DR PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
KW Hydrolase; Serine protease.
FT NON TER 32
SQ SEQUENCE 32 AA; 3330 MW; 96A3A5D1F4B19182 CRC64;

Query Match 61.0%; Score 50; DB 1; Length 32;
Best Local Similarity 50.0%; Pred. No. 0.016;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
1 IVGTAVTPGEFFYQL 16

RESULT 13
CTR2_VESCR STANDARD; PRT; 218 AA.
AC P00769; 1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin II (EC 3.4.21.1).
OS Vespa crabro (European hornet).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoecrita;
OC Aculeata; Vespidae; Vespinae; Vespa.
ON NCBI_TaxID=7445;
RN NCBI_TaxID=7445;
RP SEQUENCE.
RA Jany K.-D., Haug H.;
RT "Amino acid sequence of the chymotryptic protease II from the larvae
of the hornet, Vespa crabro.";
FBES Lett. 158:98-102(1983).
BL
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
phe-|-Xaa, Leu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: AN ADDITIONAL ARG AT THE CARBOXYL END WAS FOUND IN
SOME OF THE MOLECULES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; A00955; KYVH2C.
DR HSP; P00763; LDPO.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT ACT SITE 39 CHARGE RELAY SYSTEM.
FT ACT_SITE 84 CHARGE RELAY SYSTEM.
FT ACT_SITE 175 CHARGE RELAY SYSTEM.
FT DISULFID 25 40
FT DISULFID 148 161
FT DISULFID 171 195
SQ SEQUENCE 218 AA; 23677 MW; 509AB50DE190EB39 CRC64;

Query Match 61.0%; Score 50; DB 1; Length 218;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
```

```
Db 1 IVGGTDAPRGKYPQV 16
||||| |::|:|
1 IVGGTDAPRGKYPQV 16

RESULT 14
CAP7_PIG STANDARD; PRT; 219 AA.
AC P80015;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Azurocidin (Cationic antimicrobial protein CAP37) (Heparin-binding
protein) (HBP).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
RN NCBI_TaxID=9823;
RP SEQUENCE.
RC TISSUE=Neutrophils;
RX MEDLINE=91224149; PubMed=2026172;
RA Flodgaard H., Oestergaard E., Bayne S., Svendsen A., Thomsen J.,
Engels M., Wollmer A.;
RT "Covalent structure of two novel neutrophil leucocyte-derived
proteins of porcine and human origin. Neutrophil elastase homologues
RT with strong monocyte and fibroblast chemotactic activities.";
RL Eur. J. Biochem. 197;535-547(1991).
CC -1- FUNCTION: THIS IS A NEUTROPHIL GRANULE-DERIVED ANTIBACTERIAL AND
MONOCYTE- AND FIBROBLAST-SPECIFIC CHEMOTACTIC GLYCOPROTEIN. BINDS
HEPARIN.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
DR PIR; S15393; TRPGAZ.
DR HSP; P20160; 1A7S.
DR MEROPS; S01.971; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
KW Serine protease homolog; Glycoprotein; Chemotaxis; Antibiotic;
Heparin-binding.
FT CARBOHYD 113 113 N-LINKED (GLCNAC. .).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .).
FT DISULFID 26 42 BY SIMILARITY.
FT DISULFID 122 179 BY SIMILARITY.
FT DISULFID 152 158 BY SIMILARITY.
SQ SEQUENCE 219 AA; 24301 MW; 99129CAD88B0749D CRC64;

Query Match 59.8%; Score 49; DB 1; Length 219;
Best Local Similarity 60.0%; Pred. No. 0.18;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPM 15
||||| |::|:|
1 IVGRRAPQGEFFFL 15

RESULT 15
ACRO_RABIT STANDARD; PRT; 431 AA.
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```



```

OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Testis;
RX MEDLINE=94368861; PubMed=8086468;
RA Richardson R.T., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
RL preproacrosin-related cDNA.";
RL Biochim. Biophys. Acta 1219:215-218(1994).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY. IT
CC IS SYNTHESIZED IN A ZYMOGEN FORM, PRONACROSIN AND STORED IN THE
CC ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U05204; AAA61630.1; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.223; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16 BY SIMILARITY.
FT CHAIN 17 431 ACROSIN.
FT CHAIN 17 39 ACROSIN LIGHT CHAIN (BY SIMILARITY).
FT CHAIN 40 ? ACROSIN HEAVY CHAIN (BY SIMILARITY).
FT PROPEP ? 431 PRO-RICH.
FT DOMAIN 40 288 SERINE PROTEASE.
FT DISULFID 22 152 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 160 INTERCHAIN (BY SIMILARITY).
FT DISULFID 71 87 BY SIMILARITY.
FT DISULFID 175 244 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 234 264 BY SIMILARITY.
FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 431 AA; 46422 MW; 1C015A4E0BC0C668 CRC64;

Query Match 59.8%; Score 49; DB 1; Length 431;
Best Local Similarity 62.5%; Pred. No. 0.37;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VVGGTAAQGGFFPMV 16
Db 40 VVGQAQQAQGAFFMV 55

```

Search completed: April 11, 2003, 17:57:30
Job time : 12 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:56:10 ; Search time 16 Seconds
(without alignments)
96.134 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVGGTAAQGEFFFMV 16
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
1 number of hits satisfying chosen parameters: 283224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	259	1 TRSMG	trypsin (EC 3.4.21
2	60.5	73.8	18	2 A61577	24k serine protein
3	60	73.2	254	2 S49329	trypsin-like prote
4	55	67.1	281	2 T13596	trypsin homolog -
5	54	65.9	24	2 PN0555	alkaline trypsin-1
6	53	64.6	855	2 JC7731	membrane-bound arg
7	53	64.6	855	2 JC7775	trypsin (EC 3.4.21
8	52	63.4	20	2 A61327	chymotrypsin (EC 3
9	52	63.4	216	1 KYVH20	trypsin (EC 3.4.21
10	52	63.4	237	1 TRCV1	chymotrypsin-like
11	52	63.4	254	2 S35585	coagulation factor
12	52	63.4	274	2 I47078	coagulation trypsin V
13	52	63.4	548	2 D82175	plasma kallikrein
14	51	62.2	638	1 KQMSPL	trypsin (EC 3.4.21
15	50	61.0	32	2 S11537	chymotrypsin (EC 3
16	50	61.0	218	1 KYVH2C	alkaline trypsin-1
17	49	59.8	24	2 PN0653	azurocidin - pig
18	49	59.8	219	1 TRPAZ	acrosin (EC 3.4.21
19	49	59.8	431	2 S47538	complement C3b/C4b
20	49	59.8	613	2 S15468	coagulation factor
21	49	59.8	625	1 KFHU1	collagenolytic pro
22	48	58.5	20	2 D34817	plasma kallikrein
23	48	58.5	638	1 KQRTPL	acrosin (EC 3.4.21
24	48	58.5	638	1 KQRTPL	coagulation factor
25	47	57.3	37	2 S02176	conserved hypothet
26	47	57.3	266	2 S54146	trypsin (EC 3.4.21
27	47	57.3	416	1 KFB0	coagulation factor
28	47	57.3	819	2 B87580	factor IX - pig (f
29	46	56.1	271	2 I46580	

30	46	56.1	275	2 I46712	factor IX - rabbit
31	46	56.1	705	1 C1HURB	complement subcomp
32	45	54.9	20	2 B34817	collagenolytic pro
33	45	54.9	20	2 C34817	collagenolytic pro
34	45	54.9	42	2 B61143	trypsin (EC 3.4.21
35	45	54.9	189	2 F69473	signal sequence pe
36	45	54.9	267	1 ELHUL	leukocyte elastase
37	45	54.9	285	2 I48144	coagulation factor
38	45	54.9	452	1 A30351	coagulation factor
39	45	54.9	461	1 KFHU	coagulation factor
40	44	53.7	583	2 A29154	complement factor
41	44	53.7	1019	1 A56318	enteropeptidase (E
42	43	52.4	250	2 S55493	serine proteinase
43	43	52.4	251	1 TRHUAZ	azurocidin precurs
44	43	52.4	415	1 A34170	acrosin (EC 3.4.21
45	43	52.4	421	1 S11674	acrosin (EC 3.4.21

ALIGNMENTS

RESULT 1

TRSMG

trypsin (EC 3.4.21.4) precursor - Streptomyces griseus
C:Species: Streptomyces griseus
C>Date: 24-Apr-1984 #sequence_revision 12-May-1994 #text_change 07-May-1999
C:Accession: JQ1302; A00962
R:Kim, J.C.; Cha, S.H.; Jeong, S.T.; Oh, S.K.; Byun, S.M.
Biochem. Biophys. Res. Commun. 181, 707-713, 1991
A:Title: Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin gene
A:Reference number: JQ1302; MUID:92095977; PMID:1755852
A:Accession: JQ1302
A:Molecule type: DNA
A:Residues: 1-259 <KIM>
A:Cross-references: GB:M64471
R:Olafson, R.W.; Jurasek, L.; Carpenter, M.R.; Smillie, L.B.
Biochemistry 14, 1168-1177, 1975
A:Title: Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragment
A:Reference number: A00962; MUID:75127940; PMID:804314
A:Accession: A00962
A:Molecule type: protein
A:Residues: 37-95, 98-259 <OLA>
R:Read, R.J.; James, M.N.G.
J. Mol. Biol. 200, 523, 1988
A:Title: Refined crystal structure of Streptomyces griseus trypsin at 1.7 angstroms res
A:Reference number: A44574; MUID:86286735; PMID:3135412
A:Contents: annotation; X-ray crystallography, 1.7 angstroms
A>Note: residues 96-97 modeled as Gly-Ala
C:Genetics:
A:Gene: sprT
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-36/Domain: propeptide #status predicted <PRO>
F:37-258/Product: trypsin #status experimental <MAT>
F:37-252/Domain: trypsin homology <TRY>
F:58-74, 177-192, 204-233/disulfide bonds: #status experimental
F:73, 118, 208/Active site: His, Asp, Ser #status experimental

Query Match 100.0%; Score 82; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFFMV 16
Db 37 VVGGTAAQGEFFFMV 52

RESULT 2

A61577

24k serine proteinase (EC 3.4.21.-) - Streptomyces fradiae (fragment)
C:Species: Streptomyces fradiae

C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 06-Jan-1995

A;Accession: A61577
 R;Sinha, U.; Wolz, S.A.; Lad, P.J.
 Int. J. Biochem. 23, 979-984, 1991
 A;Title: Two new extracellular serine proteases from *Streptomyces fradiae*.
 A;Reference number: A61577; MUID:92155439; PMID:1786859
 A;Accession: A61577
 A;Molecule type: protein
 A;Residues: 1-18 <SIN>
 C;Keywords: extracellular protein; hydrolase; serine proteinase

Query Match 73.8%; Score 60.5; DB 2; Length 18;

Best Local Similarity 87.5%; Pred. No. 0.0006; Mismatches 1; Indels 1; Gaps 1;

QY 1 VVGTRAAQGEFFPMV 16

Db 1 VVGTRAAQ-EPFMV 15

RESULT 3

S49329
 trypsin-like proteinase 1 (EC 3.4.21.-) precursor - imperfect fungus (*Metarhizium anisopliae*)
 C;Species: *Metarhizium anisopliae*
 C;Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 A;Accession: JC4517; S49329
 R;Smithson, S.L.; Paterson, I.C.; Bailey, A.M.; Screen, S.E.; Hunt, B.A.; Cobb, B.D.; Co
 Gene 166, 161-165, 1995
 A;Title: Cloning and characterisation of a gene encoding a cuticle-degrading protease fr
 A;Reference number: JC4517; MUID:96105219; PMID:8529882
 A;Accession: JC4517
 A;Molecule type: DNA
 A;Residues: 1-254 <SM2>
 A;Cross-references: EMBL:X78875; MID:G556656; PIDN:CA55477.1; PID:G556657

A;Experimental source: ME1

A;Note: submitted to the EMBL Data Library, April 1994

C;Genetics:

A;Gene: try1

A;Introns: 86/3; 149/2

C;Superfamily: trypsin; trypsin homology

C;Keywords: endoplasmic reticulum; hydrolase; serine proteinase

A;Title: signal sequence #status predicted <SIG>

F;30-254/Product: trypsin-like proteinase #status predicted <PRP>

F;30-254/Product: trypsin-like proteinase #status predicted <MAT>

F;30-250/Domain: trypsin homology <TRY>

F;70,112,209/Active site: His, Asp, Ser #status: predicted

Query Match 73.2%; Score 60; DB 2; Length 254;

Best Local Similarity 68.8%; Pred. No. 0.01; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16

Db 30 IVGGSPAAAGEFFPIV 45

RESULT 4

TL13596
 trypsin homolog - fruit fly (*Drosophila melanogaster*)
 N;Alternate names: protein EG:80H7.1
 C;Species: *Drosophila melanogaster*
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 A;Accession: TL13596
 R;Benos, P.
 submitted to the EMBL Data Library, April 1999

A;Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.

A;Reference number: Z17667

A;Accession: TL13596

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-281 <BEN>

A;Cross-references: EMBL:AL031027; PIDN:CAA19843.2

C;Genetics:

A;Cross-references: FlyBase:FBgn0000481

A;Introns: 159/3

A;Note: EG:80H7.1

C;Superfamily: trypsin; trypsin homology

Query Match 67.1%; Score 55; DB 2; Length 281;

Best Local Similarity 62.5%; Pred. No. 0.082; Mismatches 10; Conservative 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16

Db 27 IVNGTTAGEFFPMV 42

RESULT 5

PN0655

alkaline trypsin-like serine proteinase (EC 3.4.21.-) F-II - earthworm (*Lumbricus rubell*

N;Alternate names: fibrinolytic enzyme

C;Species: *Lumbricus rubellus* (humus earthworm)

C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999

C;Accession: PN0655

R;Nakajima, N.; Mihara, H.; Sumi, H.

Biosci. Biotechnol. Biochem. 57, 1726-1730, 1993

A;Title: Characterization of potent fibrinolytic enzymes in earthworm, *Lumbricus rubell*

A;Reference number: PN0653; MUID:94080032; PMID:7764268

A;Accession: PN0655

A;Molecule type: protein

A;Residues: 1-24 <NAK>

C;Comment: This enzyme is a stable and potent fibrinolytic enzyme and is absorbed from

C;Keywords: hydrolase; serine proteinase

Query Match 65.9%; Score 54; DB 2; Length 24;

Best Local Similarity 56.2%; Pred. No. 0.01; Mismatches 9; Conservative 4; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16

Db 1 VVGTTNAPGEFFPQL 16

RESULT 6

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C;Species: *Rattus norvegicus* (Norway rat)

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 14-Dec-2001

C;Accession: JC7731

R;Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda,

J. Biochem. 130, 425-430, 2001

A;Title: Characterization of a membrane-bound arginine-specific serine protease from ra

A;Reference number: JC7731; MUID:21421307; PMID:11530019

A;Accession: JC7731

A;Molecule type: mRNA

A;Residues: 1-855 <KIS>

A;Cross-references: DDBJ:AB049189

A;Experimental source: strain Male, 7-week-old

C;Comment: This enzyme, localized mainly on brushborder membranes of the intestine, par

C;Keywords: protein digestion

Query Match 64.6%; Score 53; DB 2; Length 855;

Best Local Similarity 62.5%; Pred. No. 0.55; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16

Db 615 VVGTTNAGEFFPMV 630

RESULT 7

JC7775

membrane type-serine protease 1 - rat

C;Species: *Rattus norvegicus* (Norway rat)

C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C;Accession: JC7775

R;Satomi, S.; Yanasaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
 Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
 A;Title: A role for membrane-type serine protease (MT-SP1) in intestinal epithelial turn
 A;Reference number: JC7775; PMID:11573963
 A;Contents: Small intestine
 A;Accession: JC7775
 A;Molecule type: mRNA
 A;Residues: 1-855 <SAT>
 A;Cross-references: DDBJ:AB037898
 C;Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease
 ciated with epithelial migration and/or cell loss not only as an upstream activator of u
 proteins.
 C;Genetics:
 A;Gene: mt-spl
 A;Map position: basolateral cell surface

Query Match 64.4%; Score 53; DB 2; Length 855;
 Best Local Similarity 62.5%; Pred. No. 0.55; Mismatches 3; Indels 0; Gaps 0;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 DB 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 615 VVGGTNADEGEWPMQV 630

RESULT 8
 27
 C;Species: Astacus leptodactylus (narrow-fingered crayfish (fragment))
 C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
 C;Accession: A61327
 R;Zwillig, R.; Neurath, H.; Ericsson, L.H.; Enfield, D.L.
 FEBS Lett. 60, 247-249, 1975
 A;Title: The amino-terminal sequence of an invertebrate trypsin (crayfish Astacus leptod
 A;Reference number: A61327; MUID:76210806; PMID:1227963
 A;Accession: A61327
 A;Molecule type: protein
 A;Residues: 1-20 <ZWI>
 C;Comment: This enzyme could not be demonstrated as a zymogen, even in homogenates of th
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase

Query Match 63.4%; Score 52; DB 2; Length 20;
 Best Local Similarity 56.2%; Pred. No. 0.019; Mismatches 4; Indels 0; Gaps 0;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 1 IVGTDATLGEFFPYQL 16

RESULT 9
 KVVH20
 chymotrypsin (EC 3.4.21.1) II - oriental hornet
 C;Species: Vespa orientalis (oriental hornet)
 C;Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 18-Jul-1997
 C;Accession: A00954
 R;Jany, K.D.; Bekelar, K.; Pfeleiderer, G.; Ishay, J.
 Biochem. Biophys. Res. Commun. 110, 1-7, 1983
 A;Title: Amino acid sequence of an insect chymotrypsin from the larvae of the hornet, Ve
 A;Reference number: A90109; MUID:83178101; PMID:6340663
 A;Accession: A00954
 A;Molecule type: protein
 A;Residues: 1-216 <JAN>
 R;Jany, K.D.; Bekelar, K.; Ishay, J.
 Biochem. Biophys. Acta 668, 197-200, 1981
 A;Title: The amino acid sequences around the reactive serine and histidine residues of b
 A;Reference number: A90635; MUID:81208311; PMID:6786354
 A;Contents: annotation; active site
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase
 F;1-211/Domain: trypsin homology <TRY>
 F;25-40,146-159,169-193/Disulfide bonds: #status experimental

F;39,173/Active site: His, Ser #status experimental
 F;82/Active site: Asp #status predicted

Query Match 63.4%; Score 52; DB 1; Length 216;
 Best Local Similarity 50.0%; Pred. No. 0.21;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 Db 1 IVGTTNAPRGKYPYQV 16

RESULT 10
 TRCY1

trypsin (EC 3.4.21.4) I - broad-fingered crayfish
 C;Species: Astacus astacus, Astacus fluviatilis (broad-fingered crayfish)
 C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 18-Jul-1997
 C;Accession: A00951
 R;Titani, K.; Sasagawa, T.; Woodbury, R.G.; Ericsson, L.H.; Dorsam, H.; Kraemer, M.; Nei
 Biochemistry 22, 1459-1465, 1983
 A;Title: Amino acid sequence of crayfish Astacus fluviatilis trypsin I-f.
 A;Reference number: A00951; MUID:83178967; PMID:6838862
 A;Accession: A00951
 A;Molecule type: protein
 A;Residues: 1-237 <TIT>
 C;Comment: Trypsin I is one of five forms of the enzyme known to be present in crayfish
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase
 F;1-232/Domain: trypsin homology <TRY>
 F;30-46,159-174,185-213/Disulfide bonds: #status predicted
 F;45,96,189/Active site: His, Asp, Ser #status predicted

Query Match 63.4%; Score 52; DB 1; Length 237;
 Best Local Similarity 56.2%; Pred. No. 0.23;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 Db 1 IVGTDATLGEFFPYQL 16

RESULT 11

S35585
 chymotrypsin-like proteinase (EC 3.4.21.-) precursor - California red abalone
 C;Species: Haliotis rufescens (California red abalone)
 C;Date: 07-Apr-1994 #sequence_revision 19-May-1994 #text_change 22-Jun-1999
 C;Accession: S35585; S32750
 R;Groppe, J.C.; Morse, D.E.
 Arch. Biochem. Biophys. 305, 159-169, 1993
 A;Title: Molluscan chymotrypsin-like protease: structure, localization, and substrate sp
 A;Reference number: S35585; MUID:93343624; PMID:8342947
 A;Accession: S35585
 A;Molecule type: mRNA
 A;Residues: 1-254 <GRO>

A;Cross-references: EMBL:X71438; NID:G296361; PIDN:CAA50572.1; PID:G296362
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; intestine; serine proteinase; zymogen
 F;1-23/Domain: signal sequence #status predicted <SIG>
 F;23-254/Product: chymotrypsin-like proteinase #status predicted <MAT>
 F;24-251/Domain: trypsin homology <TRY>

Query Match 63.4%; Score 52; DB 2; Length 254;
 Best Local Similarity 64.3%; Pred. No. 0.24; Mismatches 3; Indels 0; Gaps 0;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 14
 :|||:|:|:|:
 Db 24 IVGGNNAAAGEFFPW 37

RESULT 12

147078
 coagulation factor IXa (EC 3.4.21.22) - sheep (fragment)

Result No.	Score	Query			Description		
		Match	Length	ID			
1	53	64.6	855	10	US-09-900-751-2	Sequence 2, Appli	
2	52	63.4	20	9	US-09-938-269-13	Sequence 13, Appl	
3	51	62.2	241	9	US-10-099-700A-4	Sequence 4, Appli	
4	51	62.2	241	9	US-10-092-004A-2	Sequence 2, Appli	
5	51	62.2	620	10	US-09-925-303-612	Sequence 1193, Ap	
6	51	62.2	855	9	US-10-099-700A-2	Sequence 2, Appli	
7	50	61.0	25	9	US-09-938-269-4	Sequence 4, Appli	
8	49	59.8	625	9	US-09-808-602-104	Sequence 104, App	
9	48	58.5	20	9	US-09-938-269-12	Sequence 12, Appl	
10	48	58.5	638	9	US-09-808-602-102	Sequence 102, App	
11	48	58.5	1128	10	US-09-888-615-97	Sequence 97, Appl	
12	46	56.1	449	10	US-09-925-303-612	Sequence 97, Appl	
13	46	56.1	688	10	US-09-874-198-7	Sequence 612, App	
14	46	56.1	688	10	US-09-874-238-7	Sequence 7, Appli	
15	46	56.1	705	9	US-09-808-602-94	Sequence 7, Appli	
16	45	54.9	18	10	US-09-861-708-2	Sequence 94, Appl	
17	45	54.9	20	9	US-09-938-269-10	Sequence 2, Appli	
18	45	54.9	20	9	US-09-938-269-11	Sequence 10, Appl	
19	45	54.9	218	10	US-09-861-708-3	Sequence 11, Appl	

; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Crayfish protease
; US-09-938-269-13

Query Match 63.4%; Score 52; DB 9; Length 20;
Best Local Similarity 56.2%; Pred. No. 0.019;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 IVGGDTLGEFFQL 16

RESULT 3
US-10-099-700A-4
; Sequence 4, Application US/10099700A
; Publication No. US2003008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-099-700A-4

Query Match 62.2%; Score 51; DB 9; Length 241;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTDADGEWPMQV 16

RESULT 4
US-10-092-004A-2
; Sequence 2, Application US/10092004A
; Publication No. US20030050251A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Araldi, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matrilysin or MTSP1
; FILE REFERENCE: 018813/0282105
; CURRENT APPLICATION NUMBER: US/10/092,004A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: PCT/US01/28137
; PRIOR FILING DATE: 2001-09-07
; CURRENT APPLICATION NUMBER: 09/657,986

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-092-004A-2

Query Match 62.2%; Score 51; DB 9; Length 241;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTDADGEWPMQV 16

RESULT 5
US-09-925-301-1193
; Sequence 1193, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1193
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (375)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (501)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (532)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (546)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-301-1193

Query Match 62.2%; Score 51; DB 10; Length 620;
Best Local Similarity 62.5%; Pred. No. 1;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 475 VVGTDADGEWPMQV 490

RESULT 6
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication No. US2003008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A

; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-099-700A-2

Query Match 62.2% Score 51; DB 9; Length 855;
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
||||| :|||:
Db 615 VVGGTDADEGEWPMQV 630

US-09-938-269-4
; Sequence 4, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Panaeus monodon tryptic
US-09-938-269-4

Query Match 61.0% Score 50; DB 9; Length 25;
Best Local Similarity 50.0%; Pred. No. 0.051;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
||||| :|||:
Db 1 VVGTAVTGGEPPYQL 16

US-09-808-602-104
; Sequence 104, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 104
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-104

Query Match 59.8% Score 49; DB 9; Length 625;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
||||| :|||:
Db 388 IVGGTASVRGEWPMQV 403

US-09-938-269-12
; Sequence 12, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Kamchatka crab
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(20)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-938-269-12

Query Match 58.5% Score 48; DB 9; Length 20;
Best Local Similarity 50.0%; Pred. No. 0.087;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFP 14
||||| :|||:
Db 1 IVGGSEATSGQEPY 14

US-09-808-602-102
; Sequence 102, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102

```
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-102

Query Match      58.5%; Score 48; DB 9; Length 638;
Best Local Similarity 50.0%; Pred. No. 3.3;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
   :|||: :|||:
Db 391 IVGGTSSWGEPWQV 406

JUL 11
US-09-888-615-97
; Sequence 97, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHITE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 97
; LENGTH: 1128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-97

Query Match      58.5%; Score 48; DB 10; Length 1128;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
   :|||: :|||:
Db 896 IVGSAAGGEPWQV 911

JUL 12
US-09-925-302-612
; Sequence 612, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 612
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-612

Query Match      56.1%; Score 46; DB 10; Length 638;
Best Local Similarity 43.8%; Pred. No. 5;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
   :|||: :|||:
Db 208 IIGGQAKMGNFPMQV 223

RESULT 13
US-09-874-198-7
; Sequence 7, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: JENSENIUS, Jens Chr.
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-198-7

Query Match      56.1%; Score 46; DB 10; Length 688;
Best Local Similarity 43.8%; Pred. No. 7.8;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16
   :|||: :|||:
Db 447 IIGGQAKMGNFPMQV 462

RESULT 14
US-09-874-238-7
; Sequence 7, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: JENSENIUS, Jens Chr.
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; TITLE OF INVENTION: USES FOR IT
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-238-7

Query Match      56.1%; Score 46; DB 10; Length 688;
Best Local Similarity 43.8%; Pred. No. 7.8;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
```

QY 1 VVGTRAAQGEFFMV 16
DB 447 IIGGQAKMGNFPPQV 462

RESULT 15

RESULT 15
US-09-808-602-94
; Sequence 94, Application US/09808602
; Patent No. US2002015115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: Macdougall, John
; TITLE OF INVENTION: No. US2002015115A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-94

Query Match	56.1%	Score 46;	DB 9;	Length 705;
Best Local Similarity	43.8%	Pred. No. 8;		
Matches	7;	Conservative	4;	Mismatches 5;
				Indels 0;
				Caps 0;

QY **1** VVGGTAAQGEEFFMV **16**

:::|::|::|:

Db **464** IIGGOKAKMGNFPPQV **479**

Search completed: April 11, 2003, 17:59:11
Job time : 15 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:56:35 ; Search time 15 Seconds
(without alignments)
31.384 Million cell updates/sec

Title: US-10-006-223-1

Perfect score: 82

Sequence: 1 VVGTRAAQGEFFFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

1 number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCITUS COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	223	1	US-08-378-091-13
2	82	100.0	223	1	US-08-483-859-13
3	82	100.0	223	1	US-08-472-173-13
4	82	100.0	223	2	US-08-487-167-13
5	82	100.0	223	2	US-08-482-816-13
6	82	100.0	223	2	US-08-296-149-13
7	82	100.0	223	2	US-08-801-499-13
8	82	100.0	223	2	US-08-615-271-13
9	82	100.0	223	3	US-09-074-660-13
10	82	100.0	223	3	US-09-074-659-13
11	82	100.0	223	3	US-09-106-468-13
12	82	100.0	223	4	US-09-106-466A-13
13	82	100.0	223	4	US-09-106-467-13
14	59	72.0	224	1	US-08-553-516-2
15	59	72.0	248	1	US-08-238-130-2
16	59	72.0	248	2	US-08-921-426-4
17	59	72.0	248	3	US-08-816-915-4
18	59	72.0	248	5	PCT-US95-07743-4
19	53	64.6	902	4	US-09-644-600-10
20	52	63.4	20	2	US-08-385-540A-15
21	52	63.4	20	2	US-08-600-273A-15
22	52	63.4	20	3	US-08-486-820-15
23	52	63.4	20	4	US-09-220-731-15
24	52	63.4	237	4	US-08-163-919A-3
25	52	63.4	237	5	PCT-US94-14073-3
26	51	62.2	855	2	US-09-027-337-2
27	51	62.2	855	4	US-09-644-600-2

28	50	61.0	25	2	US-08-385-540A-5	Sequence 5, Appli
29	50	61.0	25	2	US-08-600-273A-5	Sequence 5, Appli
30	50	61.0	25	3	US-08-486-820-5	Sequence 5, Appli
31	50	61.0	25	4	US-09-220-731-5	Sequence 5, Appli
32	50	61.0	218	4	US-09-578-303-3	Sequence 3, Appli
33	49	59.8	219	2	US-08-925-708-2	Sequence 2, Appli
34	49	59.8	238	4	US-08-944-483-64	Sequence 64, Appl
35	48	58.5	20	2	US-08-385-540A-14	Sequence 14, Appl
36	48	58.5	20	3	US-08-600-273A-14	Sequence 14, Appl
37	48	58.5	20	3	US-08-486-820-14	Sequence 14, Appl
38	48	58.5	20	4	US-09-220-731-14	Sequence 14, Appl
39	48	58.5	248	4	US-08-944-483-63	Sequence 3, Appli
40	48	58.5	638	2	US-08-681-151-3	Sequence 3, Appli
41	47	57.3	23	1	US-08-485-455D-7	Sequence 7, Appli
42	47	57.3	23	2	US-08-482-130C-7	Sequence 7, Appli
43	47	57.3	23	2	US-08-484-211C-7	Sequence 7, Appli
44	47	57.3	23	3	US-08-906-769-7	Sequence 7, Appli
45	47	57.3	23	3	US-08-906-616-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-278-091-13
; Sequence 13, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michael H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-278-091-13

Query Match 100.0%; Score 82; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFFW 16
|||||

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,167
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/296,149
 ; FILING DATE: 26-AUG-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/278,091
 ; FILING DATE: 21-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 223 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-487-167-13

Query Match 100.0%; Score 82; DB 2; Length 223;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 |||||
 Db 1 VVGTRAAQGEFFPMV 16

RESULT 5
 US-08-482-816-13
 ; Sequence 13, Application US/08482816
 ; Patent No. 5935573
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: OOMEN, Raymond P.
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/482,816
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/296,149
 ; FILING DATE: 26-AUG-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/278,091
 ; FILING DATE: 21-JUL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 223 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-482-816-13
 Query Match 100.0%; Score 82; DB 2; Length 223;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 |||||
 Db 1 VVGTRAAQGEFFPMV 16

RESULT 6
 US-08-296-149-13
 ; Sequence 13, Application US/08296149
 ; Patent No. 5939297
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: OOMEN, Raymond P.
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
 ; TITLE OF INVENTION: Reduced Protease Activity
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,149
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-390
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 223 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-296-149-13

Query Match 100.0%; Score 82; DB 2; Length 223;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 |||||
 Db 1 VVGTRAAQGEFFPMV 16

RESULT 7

APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-731 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-660-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16

RESULT 10
US-09-074-659-13
Sequence 13, Application US/09074659
Patent No. 6025342
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,659
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,167
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-659-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16

RESULT 11
US-09-106-468-13
Sequence 13, Application US/09106468
Patent No. 6114125
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,468
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-825
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-468-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16

成

```

; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-553-516-2

Query Match      72.0%; Score 59; DB 1; Length 224;
Best Local Similarity 62.5%; Pred. No. 0.019;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
Db 1 IVGTSASAGDFFIV 16

; NAME/KEY: Peptide
; LOCATION: 1..24
US-08-238-130-2

Query Match      72.0%; Score 59; DB 1; Length 248;
Best Local Similarity 62.5%; Pred. No. 0.021;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
Db 25 IVGTSASAGDFFIV 40

Search completed: April 11, 2003, 17:58:50
Job time : 17 secs

```


GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:54:35 ; Search time 35 Seconds
(without alignments)
60.915 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVCGTRAAQGEFFFMV 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

1 number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	95.1	15	21	Protease CGW-3 N-t
2	67	81.7	20	21	N-terminus of Stre
3	62	75.6	20	21	N-terminus of nove
4	61	74.4	258	23	Trichoderma harzia
5	59	72.0	248	15	Trypsin-like prote
6	59	72.0	248	16	Trypsin-like Fusar
7	59	72.0	248	17	Pre-pro-trypsin.
8	59	72.0	248	21	Pre-pro-trypsin am
9	56	68.3	360	22	Drosophila melanog
10	54	65.9	251	22	Drosophila melanog

11	54	65.9	294	22	ABB61105
12	54	65.9	296	20	AA227073
13	54	65.9	418	23	ABB06972
14	53	64.6	855	23	AAE23083
15	53	64.6	902	22	AAE98507
16	53	64.6	902	23	AAU77549
17	53	64.6	902	23	AAU80517
18	52	63.4	20	20	AA333343
19	52	63.4	20	21	AA22964
20	52	63.4	20	21	AA22964
21	52	63.4	20	21	AA22964
22	52	63.4	20	21	AA22964
23	51	62.2	241	18	AA22987
24	51	62.2	241	22	AAE06936
25	51	62.2	241	23	AAE2837
26	51	62.2	262	22	ABB58183
27	51	62.2	418	23	ABB06967
28	51	62.2	620	21	AA43748
29	51	62.2	683	21	AA19551
30	51	62.2	762	21	AA190284
31	51	62.2	851	22	ABB11428
32	51	62.2	851	22	AA225628
33	51	62.2	855	20	AA206671
34	51	62.2	855	21	AA19552
35	51	62.2	855	22	AAE06930
36	51	62.2	855	22	AAE98500
37	51	62.2	855	22	AAE3465
38	51	62.2	932	22	ABG21442
39	50	61.0	25	20	AA333333
40	50	61.0	25	21	AA22955
41	50	61.0	25	21	AA22955
42	50	61.0	25	21	AA22955
43	50	61.0	25	22	AAE07930
44	50	61.0	25	23	AAO21347
45	50	61.0	271	23	ABB62380
			418	23	ABB06971

ALIGNMENTS

RESULT 1
AAB26847
ID AAB26847 standard; peptide; 15 AA.
XX
AC AAB26847;
XX
DT 29-JAN-2001 (first entry)
XX
DE Protease CGW-3 N-terminal peptide sequence.
XX
KW Protease; CGW-3; plasminogen activator; fibrinolysis;
KW arterial thrombolysis.
XX
OS Streptomyces sp.
XX
PN CN1260394-A.
XX
PD 19-JUL-2000.
XX
PF 21-OCT-1999; 99CN-0121864.
XX
PR 21-OCT-1999; 99CN-0121864.
XX
(CHME-) CHINESE ACAD MEDICAL SCI INST BIOTECHNOL.
XX
PI Wang Y, Wu L, Chen F;
XX
DR WPI; 2000-588106/56.
XX
PT New plasmin CGW-3 and its preparation method -
XX
PS Claim 1; Page 1; 8pp; Chinese.

Drosophila melanog
L. bimastus plasmi
Guinea pig AST pro
Epithin protein.
M
Mune epithin. M
Mune type II mem
Mouse epithelin-li
Crayfish hydrolase
Crayfish protease
N-terminal of a cr
N-terminal of a cr
Crayfish protease
Human serine prote
Human membrane-ty
Human matriptase o
Drosophila melanog
Porcine AST protei
Human matriptase (
Human peptidase, H
Human membrane-ty
Human protein sequ
Tumour antigen der
Human matriptase.
Human membrane-ty
Human TADG-15. Ho
Novel human diagno
P. monodon hydroly
Tiger prawn trypti
N-terminal of a sh
N-terminal of P. m
Panaeus monodon tr
Drosophila melanog
Rabbit AST protein

CC The present invention relates to a protease termed CGW-3 which is
 CC obtained by from soil streptomyces strain C3662. CGW-3 is a serine
 CC protease and possesses fibrinolysis activity and can activate
 CC plasminogen. The N-terminal portion of CGW-3 is represented by the
 CC present sequence. The CGW-3 protease can be used in the treatment of
 CC arterial thrombosis.

XX SQ Sequence 15 AA;

Query Match 95.1%; Score 78; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VVGGTAAQGEFFPMV 15

|||||

1 VVGGTAAQGEFFPMV 15

RESULT 2

AAB03088
 AAB03088 standard; peptide; 20 AA.

AC AAB03088;

XX DT 10-OCT-2000 (first entry)

XX DE N-terminus of Streptomyces griseus trypsin.

XX KW Trypsin; bacterial; serine protease; Trichoderma; coagulant;
 XX KW hypertensive; antiinflammatory; leather preparation; silk treatment.

XX OS Streptomyces griseus.

XX JN JP2000116377-A.

XX PD 25-APR-2000.

XX PF 08-OCT-1998; 98JP-0303263.

XX PR 08-OCT-1998; 98JP-0303263.

XX (AMAN) AMANO PHARM KK.

XX WPI; 2000-369402/32.

XX A new serine protease and its preparation, used clinically in blood
 PT coagulation, hypotension and anti-inflammation -

XX Example 3; Page 7; 9pp; Japanese.

CC The invention relates to a novel serine protease from Trichoderma sp.
 CC No.9064. The N-terminus of this protease is given in AAB03085. The novel
 CC protease has trypsin-like activity, specifically cleaving peptide chains
 CC on the carboxyl side of a basic amino acid (e.g., arginine or lysine).
 CC It has a pH optimum of 7-8, and is stable between pH 5 and pH10. It has
 CC a temperature optimum of approximately 40 degrees Celsius, and is stable
 CC between 40 and 50 degrees Celsius. The invention also relates to a
 CC method for the preparation of the novel serine protease, and the use of
 CC the protease in protein degradation. The enzyme has coagulant,
 CC hypertensive and anti-inflammatory effects. It may also be used in the
 CC preparation of leather, for raw silk treatment and for the preparation of
 CC protein hydrolysate. Sequences AAB03086-B03089 represent the N-terminal
 CC 20 amino acids of trypsin from a variety of organisms.

XX SQ Sequence 20 AA;

Query Match 81.7%; Score 67; DB 21; Length 20;
 Best Local Similarity 81.2%; Pred. No. 0.00015;
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16

|||||

1 VVGGTAAQGEFFPMV 16

RESULT 3

AAB03085

ID AAB03085 standard; peptide; 20 AA.

XX AC AAB03085;

XX DT 10-OCT-2000 (first entry)

XX DE N-terminus of novel serine protease from Trichoderma sp. No.9064.

XX KW Serine protease; bacterial; trypsin activity; coagulant; hypertensive;
 XX KW antiinflammatory; leather preparation; silk treatment.

XX OS Trichoderma sp. No.9064.

XX JN JP2000116377-A.

XX PD 25-APR-2000.

XX PF 08-OCT-1998; 98JP-0303263.

XX PR 08-OCT-1998; 98JP-0303263.

XX (AMAN) AMANO PHARM KK.

XX WPI; 2000-369402/32.

XX A new serine protease and its preparation, used clinically in blood
 PT coagulation, hypotension and anti-inflammation -

XX Example 3; Page 7; 9pp; Japanese.

XX This sequence represents the N-terminal 20 amino acids of a novel serine
 CC protease from Trichoderma sp. No.9064. The novel protease has trypsin-
 CC like activity, specifically cleaving peptide chains on the carboxyl side
 CC of a basic amino acid (e.g., arginine or lysine). It has a pH optimum of
 CC 7-8, and is stable between pH 5 and pH10. It has a temperature optimum of
 CC approximately 40 degrees Celsius, and is stable between 40 and 50 degrees
 CC Celsius. The invention also relates to a method for the preparation of
 CC the novel serine protease, and the use of the protease in protein
 CC degradation. The enzyme has coagulant, hypertensive and anti-inflammatory
 CC effects. It may also be used in the preparation of leather, for raw silk
 CC treatment and for the preparation of protein hydrolysate.

XX SQ Sequence 20 AA;

Query Match 75.6%; Score 62; DB 21; Length 20;

Best Local Similarity 68.8%; Pred. No. 0.0011;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16

|||||

1 VVGGTAAQGEFFPMV 16

RESULT 4

ABB83181

ID ABB83181 standard; Protein; 258 AA.

XX AC ABB83181;

XX DT 16-AUG-2002 (first entry)

XX DE Trichoderma harzianum Pral protease.

XX KW Pral; protease; enzyme; fungicide; insecticide; serine-peptidase.

XX OS Trichoderma harzianum.

XX Key Location/Qualifiers

FT Misc-difference 236

FT XX /note= "Encoded by GAC"

PN XX WO200244359-A1.

PD XX 06-JUN-2002.

PF XX 30-NOV-2001; 2001WO-ES00471.

PR XX 01-DEC-2000; 2000ES-0002897.

XX XX (NEWB-) NEWBIOTECHNIC SA.

PA (UYSE-) UNIV SEVILLA.

PA (UYSA-) UNIV SALAMANCA.

XX XX

PI Suarez Fernandez B, Rey Barrera M, Monte Vazquez E;

FI Llobell Gonzalez A;

XX XX

DR WPI; 2002-471830/50.

DB N-PSDB; ABN83394.

XX XX

XX New proteolytic enzyme from *Trichoderma harzianum*, useful e.g. for protecting plants against fungal attack, also related nucleic acid -

PS Claim 1; Page 44-46; 51pp; Spanish.

XX XX

CC The present sequence is the protein sequence for Pral protease from *Trichoderma harzianum*. Pral, a serine-peptidase, catalyses irreversible inactivation of enzymes and proteins essential for pathogenicity of fungi. Pral can be used to degrade proteins and peptides, especially structural components of the cell walls of fungi, insects and arachnids. Particularly, Pral can be used (optionally in combination with chemical fungicides) to protect plants, animals, harvested crops and foods against fungi; to generate protoplasts and yeast extracts; for recovery of manoproteins; in preparation of wine and (grape) juice; for removal of dental plaque; in tooth-, denture- and contact lens-cleaning solutions; to remove biofilms; to treat or clean textiles; as disinfectant and to prevent contamination of analytical samples.

XX XX

XX Sequence 258 AA;

Query Match 74.4%; Score 61; DB 23; Length 258;

Best Local Similarity 68.8%; Pred. No. 0.026;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAAGGEPFPMV 16

DB :|||||:|||||:

30 IVGGTTAAUGGEPFIV 45

XX XX

XX RESULT 5

XX AAR66999

ID AAR66999 standard; Protein; 248 AA.

XX XX

XX AAR66999;

12-JUN-1995 (first entry)

Trypsin-like protease.

Trypsin; protease; detergent; surfactant.

Fusarium oxysporum DSM 2672.

Key Location/Qualifiers

FT Peptide 1..23

FT /label= Pre-propeptide

XX XX

PN WO9425583-A.

XX XX

PD 10-NOV-1994.

XX XX

PF 04-MAY-1994; 94WO-DK00177.

XX XX

PR 05-MAY-1993; 93DK-0000523.

XX XX (NOVO) NOVO-NORDISK AS.

PA XX

PI Branner S, Hastrup S;

XX XX

DR WPI; 1994-358261/44.

DR N-PSDB; AAQ74774.

XX XX

PT Recombinant trypsin-like protease - useful as e.g. additive enzyme in detergent compsn.

FT enzyme in detergent compsn.

XX XX

PS Disclosure; Page 33-34; 44pp; English.

XX XX

CC An active recombinant trypsin-like protease comprises residues 25-224 of the protein given in AAR66999. cDNA encoding the protease was isolated from a cDNA library of *F. oxysporum* DSM 2682. Expression in *Bacillus*, *Streptomyces*, *Escherichia coli*, *Saccharomyces*, *Aspergillus* or *Fusarium* hosts, with addition of *Bacillus metallo* protease to the fermentation broth, yielded the mature protease.

XX XX

SQ Sequence 248 AA;

Query Match 72.0%; Score 59; DB 15; Length 248;

Best Local Similarity 62.5%; Pred. No. 0.055;

Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAAGGEPFPMV 16

DB :|||||:|||||:

25 IVGGTSASAGDFPFI 40

XX XX

XX RESULT 6

XX AAR84716

ID AAR84716 standard; Protein; 248 AA.

XX XX

XX AAR84716;

19-MAR-1996 (first entry)

Trypsin-like *Fusarium* protease.

Trypsin-like protease; detergent; surfactant; enzyme engineering; enzyme stabilisation.

Fusarium oxysporum DSM 2672.

Key Location/Qualifiers

FT Peptide 1..6

FT /label= Pro-peptide

FT Peptide 8..24

FT /label= Sig_peptide

FT Region 32..36

FT /label= Loop-I

FT Region 44..47

FT /label= Loop-II

FT Region 55..58

FT /label= Loop-III

FT Region 67..75

FT /label= Loop-IV

FT Region 81..88

FT /label= Loop-V

FT Region 99..105

FT /label= Loop-VI

FT Region 116..124

FT /label= Loop-VII

FT Region 130..137

FT /label= Loop-VIII

FT Region 151..158

FT /label= Loop-IX

FT Region 170..186

FT /label= Loop-X

FT Region 191..196

FT FT /label= Loop-XI
FT 210..214
FT /label= Loop-XI
FT 220..230
FT /label= Loop-XII
XX WO9530743-A1.
FN
XX
XX
XX 16-NOV-1995.
XX
XX 04-MAY-1995; 95WO-DK00180.
XX
XX 04-MAY-1994; 94DK-0000509.
PR (NOVO) NOVO-NORDISK AS.
PI Fredholm H, Von Der Osten C;
XX
DR WPI; 1995-404110/51.
N-PSDB; AAT05182.
Variants of a parent trypsin-like Fusarium protease - used as
constituents in detergent compositions
Claim 1; Page 57-58; 80pp; English.
XX The trypsin-like protease (AAR84716) from Fusarium oxysporum DSM 2672
XX includes 13 surface loop regions which can be substituted, esp. by
XX corresponding loop regions II and IV of cattle trypsin 2ptn, or
XX loops IV or VI of rat trypsin 1trm, or by other modified loop
XX structures. The protease variants are obtd. by site-directed
XX mutagenesis of encoding cDNA (AAT05182). They show improved
XX proteolytic activity and decreased susceptibility to oxidation than
XX the parent enzyme, and are useful in detergent compositions and
XX additives.
XX Sequence 248 AA;
XX
XX Query Match 72.0%; Score 59; DB 16; Length 248;
XX Best Local Similarity 62.5%; Pred. No. 0.055; 2; Indels 0; Gaps 0;
XX Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 1 VVGGTAAAGGEPFPMV 16
25 IVGGTSASAGDPFFIV 40
RESULT 7
AAR8470
AAR88470 standard; Protein; 248 AA.
XX
XX AAR88470;
XX
XX 03-APR-1996 (first entry)
XX Pre-pro-trypsin.
XX Trypsin; alkaline protease; SP387; host cell; Fusarium graminearum;
XX heterologous gene expression.
XX Fusarium oxysporum.
XX
XX Key Location/Qualifiers
XX Peptide 1..24
XX /label= Pre-propeptide
XX WO9600787-A1.
XX
XX 11-JAN-1996.
XX
XX 15-JUN-1995; 95WO-US07743.
XX
XX 15-MAR-1995; 95US-0404678.
PR

PR 30-JUN-1994; 94US-0269449.
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX
XX Moyer DL, Royer JC, Shuster JR, Yoder W;
XX
XX WPI; 1996-077498/08.
XX N-PSDB; AAT10181.
XX
XX Non-toxic, non-toxicogenic, non-pathogenic recombinant Fusarium host
XX cell - used to produce heterologous proteins, pref. enzymes,
XX hormones, growth factors or receptors
XX
XX Claim 11; Page 20; 38pp; English.
XX
XX Fusarium oxysporum SP387 trypsin-like protease (AAR88470) is produced
XX in a novel, non-toxic, non-toxicogenic, non-pathogenic,
XX protease-deficient recombinant host, pref. Fusarium graminearum
XX ATCC 20334, by expression of the encoding gene (AAT10181) operably
XX linked to promoter and terminator sequences (AAT10184-85). Recombinant
XX enzyme is obtd. in yields of at least 0.5 g/l host cell.
XX
XX Sequence 248 AA;
XX
XX Query Match 72.0%; Score 59; DB 17; Length 248;
XX Best Local Similarity 62.5%; Pred. No. 0.055; 4; Mismatches 2; Indels 0; Gaps 0;
XX Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 VVGGTAAAGGEPFPMV 16
25 IVGGTSASAGDPFFIV 40
DB
RESULT 8
AAR803659
ID AAR803659 standard; Protein; 248 AA.
XX
XX AAR803659;
XX
XX 25-SEP-2000 (first entry)
XX Pre-pro-trypsin amino acid sequence.
XX
XX Non-toxic; non-pathogenic; recombinant protein production; protease;
XX pre-pro-trypsin.
XX
XX Fusarium oxysporum.
XX
XX US6060305-A.
XX
XX 09-MAY-2000.
XX
XX 13-MAR-1997; 97US-0816915.
XX
XX 30-JUN-1994; 94US-0269449.
XX 15-MAR-1995; 95US-0404678.
XX 04-OCT-1996; 96US-0726105.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX
XX Wendy YT, Shuster JR, Moyer DL, Royer JC;
XX
XX WPI; 2000-349678/30.
XX N-PSDB; AAA53331.
XX
XX New non-pathogenic recombinant fusarium host cell, useful for
XX expressing heterologous proteins especially fungal enzymes such as
XX alkaline endoglucanase or alkaline protease -
XX
XX Disclosure; Column 15-18; 32pp; English.
XX
XX The invention relates to a non-toxic, non-pathogenic recombinant Fusarium
XX host cell of the section Discolor, with ATCC accession number 20334. The
XX

CC cell is used in the recombinant production of proteins. The present
 CC sequence represents the F. oxysporum pre-pro-trypsin protein sequence,
 CC which is an example of an enzyme which can be expressed by the cells of
 CC the invention. The cells are useful for expressing heterologous proteins
 CC especially fungal enzymes such as alkaline endoglucanase or alkaline
 CC proteases, e.g. F. oxysporum pre-pro trypsin gene, and also hormones,
 CC growth factors and receptors. The cells are non-toxic and are efficient
 CC in the recombinant production of fungal enzymes.

XX SQ Sequence 248 AA;

Query Match 72.0%; Score 59; DB 21; Length 248;
 Best Local Similarity 62.5%; Pred. No. 0.055;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:|:
 Db 25 IVGTSASAGDFPIV 40

LT 9

AB60104
 ID ABB60104 standard; Protein; 360 AA.
 XX
 AC ABB60104;

26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 7104.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04207.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 7104; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175), and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 360 AA;

Query Match 68.3%; Score 56; DB 22; Length 360;
 Best Local Similarity 66.7%; Pred. No. 0.27;

Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGTRAAQGEFFPM 15
 :|||:|:|:|:|:
 Db 102 IVGTRASGKEFFPM 116

RESULT 10

ABB64578
 ID ABB64578 standard; Protein; 251 AA.

XX ABB64578;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 20526.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08681.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 20526; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175), and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 251 AA;

Query Match 65.9%; Score 54; DB 22; Length 251;
 Best Local Similarity 50.0%; Pred. No. 0.39;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:|:
 Db 28 IVGTRAREGQPHQI 43

RESULT 11

ABB61105
 ID ABB61105 standard; Protein; 294 AA.

XX ABB61105;

XX

DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 10107.
 DE
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmacological.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 Venter JC, Adams M, Li PWD, Myers EW;
 WPI; 2001-656860/75.
 N-PSDB; ABL05208.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 10107; 21bp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 294 AA;
 Query Match 65.9%; Score 54; DB 22; Length 294;
 Best Local Similarity 66.7%; Pred. No. 0.46;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 1 VVGGTAAAGGFFPFM 15
 :|||:|||||
 37 IVGAKAAGREFFPM 51
 RESULT 12
 AAY27073
 ID AAY27073 standard; Protein; 296 AA.
 XX
 AC AAY27073;
 XX
 DT 18-OCT-1999 (first entry)
 DE L. bimastus plasmin mature protein sequence.
 XX
 KW Lumbricus bimastus; earthworm; plasmin gene.
 XX
 OS Lumbricus bimastus.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 4 /note= "encoded by GC"
 FT Misc-difference 12 /note= "encoded by TGC"
 FT

FT Misc-difference 243 /note= "encoded by TAG"
 FT Misc-difference 268 /note= "encoded by TAA"
 FT Misc-difference 274 /note= "encoded by TAA"
 FT Misc-difference 279 /note= "encoded by TAA"
 FT Misc-difference 280 /note= "encoded by TAA"
 FT Misc-difference 292 /note= "encoded by TAA"
 FT Misc-difference 296 /note= "encoded by TAG"
 XX
 PN CN1208770-A.
 XX
 PD 24-FEB-1999.
 XX
 PF 11-JUN-1998; 98CN-0102257.
 XX
 PR 11-JUN-1998; 98CN-0102257.
 PA (VIRO-) VIROLOGY RES INST CHINA PREVENTIVE MEDIC.
 XX
 PI Fu S, Liang G, Meng X;
 XX WPI; 1999-313741/27.
 DR N-PSDB; AAX89869.
 XX
 XX Lumbrical fibrinolysin gene nucleotide series, and method for clone
 PT of same - is composed of 88 nucleotides in which position 1-726
 PT nucleotide is gene matured peptide sequence, position 72-729
 PT nucleotide is terminal codon TAG
 XX
 PS Claim 3; Page 1; 9pp; Chinese.
 XX
 CC The invention provides a Lumbricus bimastus (a kind of earthworm) plasmin
 CC gene. The gene is composed of 88 nucleotides in which position 1-726
 CC nucleotide is gene matured peptide sequence, position 72-729 nucleotide
 CC is terminal codon TAG. The expressed protein from the gene possesses
 CC tremendous medical prospect. The present sequence represents the
 CC L. bimastus plasmin gene mature protein sequence.
 XX
 SQ Sequence 296 AA;
 Query Match 65.9%; Score 54; DB 20; Length 296;
 Best Local Similarity 56.2%; Pred. No. 0.47;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVGGTAAAGGFFPFM 16
 :|||:|||||
 Db 1 VVGGTAAAGGFFPFM 16
 RESULT 13
 ABB06972
 ID ABB06972 standard; Protein; 418 AA.
 XX
 AC ABB06972;
 XX
 DT 20-JUN-2002 (first entry)
 DE Guinea pig AST protein sequence SEQ ID NO:37.
 XX
 KW Airway specific trypsin-like protease; AST; protease activated receptor;
 KW PAR; mucus production promotion; cell proliferation; calcium flow;
 KW EGFR pathway activation; epidermal growth factor receptor; enzyme.
 XX
 OS Cavia porcellus.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..186

FT Protein /label= propeptide
 FT 187..418
 XX /label= trypsin-like protein
 PN WC000218562-A1.
 XX
 XX 07-MAR-2002.
 XX
 XX 28-AUG-2001; 2001WO-JP07349.
 XX
 XX 28-AUG-2000; 2000JP-0257104.
 PR
 PR 05-MAR-2001; 2001JP-0059753.
 XX
 XX (TEIJ) TEIJIN LTD.
 PA
 XX
 XX Eguchi H, Chokki M, Yamamura S, Mita R, Masegi T;
 XX WPI; 2002-315539/35.
 DR N-PSDB; ABL50743.
 DR
 DR
 FT Airway-specific trypsin-like enzymes for use in diagnosis and screening
 XX compounds or polypeptides as inhibitors of AST activity, PAR activation
 XX and mucus production, and judging therapeutic efficacy
 PS Example 10; Page 146-148; 165pp; Japanese.
 The present invention describes an airway-specific trypsin-like enzyme
 (AST) is a protein comprising the whole or a part of the amino acid
 sequence of (AB06964 (I)), in which a propeptide moiety containing all
 or part of an amino acid sequence of AST between Met at position 1 and
 Arg at position 186 is bonded to a trypsin-like moiety containing Ile at
 position 187 to Ile at position 418 in a 533 amino acid sequence via a
 disulfide bond. The AST enzyme can be used in the diagnosis and screening
 of compounds and polypeptides as inhibitors of AST activity, protease
 activated receptor (PAR) activation, mucus production promotion, cell
 proliferation, calcium flow into cells or EGFR pathway activation by AST
 and judging therapeutic efficacy. The present sequence represents Guinea
 pig AST from the present invention.
 XX
 XX Sequence 418 AA;
 SQ
 Query Match 65.9%; Score 54; DB 23; Length 418;
 Best Local Similarity 62.5%; Pred. No. 0.69;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGGTAAQGEFFPMV 16
 ||||| :||| :
 187 VVGGTQADQGDWPMQV 202
 RESULT 14
 AAE23083
 ID AAE23083 standard; Protein; 855 AA.
 XX
 XX AAE23083;
 DT 21-AUG-2002 (first entry)
 XX
 XX Epithin protein.
 XX
 KW Transgenic; transgenic animal; pharmacological therapy; gene therapy;
 KW phenotype modulation; genetic disease; epithin protein.
 XX Unidentified.
 OS
 XX WO200203787-A2.
 XX
 XX 17-JAN-2002.
 XX
 XX 06-JUL-2001; 2001WO-US21427.
 XX
 XX 06-JUL-2000; 2000US-216109P.
 PR
 PR 06-JUL-2000; 2000US-216251P.
 PR

PR 06-JUL-2000; 2000US-216258P.
 PR 06-JUL-2000; 2000US-216768P.
 PR 10-JUL-2000; 2000US-217449P.
 PR 10-JUL-2000; 2000US-217450P.
 PR 10-JUL-2000; 2000US-217660P.
 PR 27-JUL-2000; 2000US-221491P.
 PR 27-JUL-2000; 2000US-221669P.
 PR 27-JUL-2000; 2000US-221670P.
 PR 07-AUG-2000; 2000US-223170P.
 PR 07-AUG-2000; 2000US-223172P.
 PR 07-AUG-2000; 2000US-223460P.
 PR 26-OCT-2000; 2000US-244037P.
 PR 26-OCT-2000; 2000US-244111P.
 PR 26-JUN-2001; 2001US-301217P.
 XX
 XX (DELT-) DELTAGEN INC.
 XX
 XX Allen KD, Leviten MW;
 PI
 XX WPI; 2002-154853/20.
 DR N-PSDB; AAD37039.
 DR
 XX Novel non-human transgenic animal, preferably transgenic mice
 PT comprising disruption in target gene, e.g., trypsin gene, useful for
 PT identifying an agent that modulates expression or function of target
 PT gene
 XX
 XX Disclosure; Fig 6; 74pp; English.
 PS
 XX The present invention relates to non-human transgenic animals preferably
 CC transgenic mice comprising disruption in target gene such as trypsin
 CC gene. The invention also relates to compositions and methods relating
 CC to the characterization of gene functions. The transgenic animals are
 CC useful for identifying an agent that modulates the expression or function
 CC of a target. They are useful for identifying an agent that modulates a
 CC phenotype associated with a disruption in trypsin genes or limulus
 CC clotting factor protease-like genes by administering an agent to the
 CC transgenic animal and determining whether the agent modulates the
 CC phenotype where the agent has effect on decreased body weight, decreased
 CC thymus weight, decreased thymus to body weight ratio, increased pre-pulse
 CC inhibition, significant decrease in their response latency to the hot
 CC plate test or a decreased response threshold to metrazol. Agents that
 CC modulate the expression, function or activity of the target gene are
 CC useful for treating a disorder associated with a mutation in trypsin
 CC gene or in limulus clotting factor protease-like gene. The transgenic
 CC animals are useful for testing the efficacy of proposed genetic and
 CC pharmacological therapies for human genetic diseases. They are useful
 CC as models for diseases, disorders or conditions associated with
 CC phenotypes relating to a disruption in a target and to identify drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating a disease or other phenotypic characteristics of the animal.
 CC The present sequence is, epithin protein. This sequence is used in the
 CC exemplification of the invention.
 XX
 XX Sequence 855 AA;
 SQ
 Query Match 64.6%; Score 53; DB 23; Length 855;
 Best Local Similarity 62.5%; Pred. No. 2.2;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVGGTAAQGEFFPMV 16
 ||||| :||| :
 Db 615 VVGGTNADGEWPMQV 630
 RESULT 15
 AAB98507
 ID AAB98507 standard; Protein; 902 AA.
 XX
 XX AAB98507;
 XX
 XX 03-AUG-2001 (first entry)
 DT
 XX

DE Murine epithin.
XX
XX Murine; TADG-15; cytostatic; vaccine; ovarian tumour; cancer; epithin;
KW tumour antigen-derived gene 15; serine protease.
XX
XX OS Mus musculus.
XX
XX WQ200129056-A1.
XX
XX PD 26-APR-2001.
XX
XX PF 20-OCT-2000; 2000WO-US29095.
XX
XX PF 20-OCT-1999; 99US-0421213.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX O'Brien TJ, Tanimoto H;
XX
XX WPI; 2001-381031/40.

Novel extracellular serine protease, termed tumor antigen-derived gene 15 protein overexpressed in carcinomas and DNA encoding it, for diagnosis, treatment, prevention of cancer, particularly breast, ovarian cancer -

Disclosure; Page 97-99; 130pp; English.

The present invention relates to human tumour antigen-derived gene 15 (TADG-15) protein and coding sequence (see AAB23601 and AAB98500). TADG-15 is an extracellular serine protease. It was found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein or its fragments of 9-20 residues that lack TADG-15 protease activity are useful for vaccinating an individual against TADG-15, having, suspected of having or at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a diagnostic or therapeutic target in cancer. The present sequence was used in a sequence homology alignment with TADG-15.

XX
XX
SQ Sequence 902 AA;

Query Match 64.6%; Score 53; DB 22; Length 902;
st Local Similarity 62.5%; Pred. No. 2.4;
tches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGCTRAAQGEFPFMV 16
Db 615 VVGTTNADEGEPMQV 630

Job completed: April 11, 2003, 17:57:13
Job time : 37 secs